Vision, Modeling, and Visualization (2019)
H.-J. Schulz, M. Teschner, and M. Wimmer (Eds.)

Local Remote Photoplethysmography Signal Analysis for Application in Presentation Attack Detection

Benjamin Kossack[1], Eric L. Wisotzky[1,2], Anna Hilsmann[1] and Peter Eisert[1,2]

[1]Fraunhofer HHI, Computer Vision and Graphics, Berlin, Germany
[2]Humboldt-Universität zu Berlin, Visual Computing, Berlin, Germany

Abstract

This paper presents a method to analyze and visualize the local blood flow through human skin tissue within the face and neck. The method is based on the local signal characteristics and extracts and analyses the local propagation of blood flow from video recordings. In a first step, the global pulse rate is identified in RGB images using normalized green color channel intensities. We then calculate for an image sequence, a local remote photoplethysmography (rPPG) signal that is presented by a chrominance-based signal. This local rPPG signal is analyzed and then used to extract the local blood flow propagation from signal-to-noise ratio (SNR) and pulse transit time (PTT) maps. These maps are used to visualize the propagation of the blood flow (PTT) and reveal the signal quality of each spatial position (SNR). We further proved a novel pulse rate based skin segmentation method, that is based on the global pulse rate and the statistical properties of the SNR map. This skin segmentation method allows a direct application in liveliness detection, e.g., for presentation attack detection (PAD). Based on the described local blood flow analysis, we propose a PAD system, that specializes in identifying a partial face and neck coverage in the video. The system is tested using datasets showing a person with different facial coverings, such as a mask or a thick layer of makeup. All tested masks can be detected and identified as presentation attacks.

1. Introduction

The human face is an important source of information about a human being, such as its condition, e.g., measured by the pulse rate, and is therefore used in a variety of different other applications. Particularly relevant fields of application for these techniques are medical or security technology. An optical measuring technique called photoplethysmography (PPG) is commonly used to measure the human pulse rate. The principle of PPG is based on human blood circulation and the fact that blood absorbs more light than surrounding tissue, so variations in blood volume affect light transmission or reflectance accordingly [TMSY14]. A PPG sensor placed directly on the skin optically detects these changes in blood volume [TMSY14]. Wearing this sensor can cause patients (especially infants and children) to feel uncomfortable and nervous, which can have a negative effect on the measurement. To overcome this, remote photoplethysmography (rPPG) allowing contactless measurements of the pulse rate with a regular camera has been developed [ZTWM18]. This technology has become more attractive in recent years and is also based on human blood circulation [ZTE*18]. In addition to the measurement of vital signs, rPPG is also used in security applications to detect presentation attacks [SL16, HM19, LKZ*17, NSV17].

The heart generates a blood volume pulse (BVP) with each beat, which is the source of the blood circulation. The resulting blood flow through the circulatory system leads to a continuous change in skin color. This effect can be observed more strongly when a person is physically active, e.g., after climbing stairs. The heart rate rises and the face color is getting redder. However, this constant color change is mostly imperceptible for the human eye. With rPPG techniques, this color variation is detected from a video, and the pulse rate can be determined. Wu et al. [WRS*12] presented a method to magnify these imperceptible color changes to visualize these for the human eye in video recordings. This method is noted as Eulerian Video Magnification (EVM).

The majority of rPPG-related literature applies the method to extract vital signs globally [ZTWM18]. In [RWAH17, PMP10, TL15, DJ13] the pulse rate is estimated by extracting and analyzing the subtle color changes in the skin area. In addition to a pulse rate or other vital signs, videos provide even more characteristics. The local extraction of these characteristics allows a more robust and differentiated recognition of the investigated properties. Our work directs at the local (i.e., spatial) analysis of the pulse rate and blood flow through human skin tissue from videos with the aim to visualize the blood flow.

In this paper, we contribute a new approach to the analysis and visualization of local blood flow based on a chrominance-based rPPG signal. In addition, we present a novel method for segmentation of living skin tissue in the face and neck. This segmentation

DOI: 10.2312/vmv.20191327

relies on the pulse rate of the recorded subject and the signal-to-noise ratio (SNR) of the local rPPG signal.

Based on this analysis and segmentation, we propose a new presentation attack detection (PAD) system, which makes it possible to debunk presentation attacks with a lifeless mask or high-resolution image. The proposed system makes it possible to detect persons wearing partial masks or thick makeup.

The paper is organized as follows: the next section presents prior work on heart rate measurement using rPPG, spatial rPPG analysis, and on face PAD. Then, the proposed local blood flow analysis and visualization approach, is described in detail in Section 2, together with the performed data acquisition. We discuss experiments and results in Section 3. Lastly, a conclusion is drawn, and proposals for future research are made in the last section.

1.1. Heart Rate Measurement

The extraction of human vital signs from video recordings of the face is an emerging topic that has grown rapidly in recent years and has produced a variety of publications.

Poh et al. [PMP10] present a non-contact, automated, and motion tolerant heart rate measurement from video images based on blind source separation (i.e., independent component analysis (ICA)). Rapczynski et al. [RWAH17] calculate the rPPG signal from the green channel of the subjects skin pixels. In [TL15], different color spaces are compared to find the best-suited space for human pulse calculation. It is shown that the hue component in the HSV color space delivers the most accurate pulse rate measurement. De Haan and Jeanne [DJ13] present an algorithm to calculate a chrominance-based rPPG signal. This algorithm works robustly regardless of the subjects skin tone and the color of the illumination with colored light sources (i.e., for nonwhite illumination).

1.2. Spatial rPPG Analyses

The two-dimensional properties of the rPPG signal are analyzed in [Yan15] and [ZTE*18] and presented using different types of visualization maps as amplitude, velocity, or SNR map. In [Yan15], a method for estimating blood flow velocities from videos of human faces is presented. The velocity is calculated from the relative phase shift of the frequency component corresponding to the heart rate in the frequency domain. It is assumed that the difference between neighboring phase values directly corresponds to the velocity at this point. Applying the 2D Sobel operator to the calculated phase shifts results in the desired velocity map.

Zaunseder et al. [ZTE*18] introduced a contactless method for estimating perfusion speed from videos. The spatially separated rPPG signals are bandpass filtered based on the heart rate measured with an electrocardiogram (ECG). After defining a region of interest (ROI) on the subject's forehead, a point *P0* inside this ROI is selected. The distance to a chosen point *P1* is calculated by using the Euclidean distance. The time delay between these specific positions can be extracted from the filtered rPPG signals of these positions. Then, an estimate of the pulse wave perfusion speed can be calculated with the calculated distance and time delay. Since the calculation is performed on the pixel values, the result is neither a physical speed nor can it be transferred to other data.

1.3. Presentation Attack Detection

The use of a facial recognition system for authentication has become widespread. Biometric authentication systems based on facial recognition are already used in border security systems and to unlock smartphones. Although widely used and highly accurate, facial recognition algorithms suffer from vulnerability to simple spoofing attacks. An attack on face recognition based security systems are stated as biometric presentation attack. Such an attack is the attempt to bypass a biometric security system by impersonating a target victim holding the desired authorization [MSC16]. During such presentation attacks, the security system may not be able to distinguish between the biological trait of the authorized person and the presented object.

This work is focused on the support of facial recognition security systems. These systems are especially endangered because the face of a person is not private. In our society, it is nearly impossible to keep the face private and avoid recordings of it. Potential attackers could use a digital camera or social network content to grab the face of an authorized person. This data can be used to print a high-quality image (photo attacks) or to create a realistic 3D mask of the authorized person (mask attacks) with which the facial recognition security system is then attacked [GMF14].

It is challenging to detect mask attacks because they imitate entire 3D structures and facial colors. A promising approach is to measure signals of a living body which are attributes to the nervous system (e.g., pulse rate) [GMF14]. By applying rPPG methods to a face, it is possible to detect whether the whole face is covered or not. In [LKZ*17], a robust anti-spoofing method is proposed that is based on the fact that a pulse signal exists in a real living face but not in any mask or printed material. Xi et al. define for their method the lower half of the face, including cheeks, nose, mouth, and chin as an ROI. After applying time filtering and FFT, a six-dimensional feature vector is constructed as input for a support vector machine to determine whether a presentation attack occurs.

Heusch et al. [HM19] propose new features containing long term spectral statistics of pulse signals obtained through rPPG. In a wide variety of experiments, they successfully apply their approach to face presentation attack detection.

For PAD application, the pulse rate is mostly extracted from the whole face [SL16, HM19], or specific regions [LKZ*17, NSV17]. These methods are not able to determine whether partial areas of the face are covered because they calculate the rPPG signal over all pixels of their defined ROI. In these cases, partial face coverage would reduce signal quality, but there would still be a measurable rPPG signal. Thus, the recording of a person with partial masking would not be detected as a presentation attack.

2. Proposed Framework

2.1. Global Pulse Rate Estimation

We present a framework that makes it possible to analyze and visualize the blood flow through the human face and neck using RGB video recordings. The output of our framework is two maps, an SNR map, and a pulse transit time (PTT) map. The SNR map illustrates the signal quality of the corresponding part within the face

and neck. The PTT map visualizes which spatial position the BVP reaches at which point in time. With the PTT map, a blood flow path can be identified. We assume that the recorded subject is facing the camera throughout the whole recording. The framework is always applied to image sequences with a length of 10 s. In order to compensate for the movement of the subject, the images of the recording are spatially aligned. We adjust the rotation and translation between successive images via the normalized cross-correlation. Thus the subject is registered image by image.

To analyze and visualize the human blood flow, we are applying the framework shown in Figure 1. As in [WRS*12], we filter each image with a low-pass (e.g., with a Gaussian or average filter) and downscale it afterward. This filter and downscale step is repeated until the desired image size and/or signal quality is reached (for our data, two to three times). The spatial filtering increases the SNR and reduces the computational cost through downscaling of the images.

After spatial filtering, the pulse rate of the recorded subject is calculated. In [KKLP15], it is shown that the forehead and both cheeks are good candidates for a computationally efficient ROI. Therefore, we calculate the global pulse rate f_{pr} from the forehead region using the normalized green color channel intensities [RWAH17]. The green channel contains the strongest PPG signal, based on the fact that hemoglobin absorbs green light better than red and penetrates deeper into the skin than blue [VSN08].

We determine a rectangular region on the forehead based on the eye position, whereby the coordinates of the eyes are detected using Haar feature-based cascade classifiers as proposed by Viola and Jones [VJ01]. Figure 2 illustrates the determination of the forehead region. We calculate the distance X between the center of both eyes from the return values of the cascade classifier. As x-coordinates for the left and right border of the ROI, we use the x-coordinates of the corresponding eye centers. The y-coordinate of the center of the ROI is set $(X/2)$ above the y-coordinate of the right eye. The total height of the ROI is $(X/2)$.

On the basis of [RWAH17], the rPPG signal of the forehead region is calculated. For each frame, the mean of all normalized green value G_n is calculated by:

$$G_n = \frac{1}{N}\sum_{i=1}^{N}\frac{g_i}{r_i + g_i + b_i}, \tag{1}$$

where r_i, b_i and g_i are the intensity values for each pixel i of the red, blue and green color channel of the frame, and N is the total amount of pixels within the forehead region.

To focus on the physiologically possible pulse rate range, the signal $G_n(t)$ is bandpass filtered. The applied filter is a zero-phase-filter with fifth-order Butterworth bandpass coefficients and a passband range between 0.5 Hz and 3.3 Hz (corresponding to 30 beats per minute (BPM) and 200 BPM) [RWAH17]. $G_{nBP}(t)$ represents the bandpass filtered signal. Figure 2 shows the waveforms in the time domain of the calculated rPPG signal G_{nBP} (1) and the reference ECG (3), measured simultaneously with the video recording. To extract the pulse rate from these waveforms, we apply an FFT and calculate the magnitude (see Figure 2 plot (2) and (4)). The frequency component with the highest magnitude represents the global pulse rate f_{pr}. In the example shown in Figure 2, the frequency components with the highest magnitude for the calculated G_{nBP} (2) and the measured ECG (4) are identical. In general, the method of [RWAH17] leads to precise and satisfactory results for all recordings of our dataset (see Section 2.5).

The local blood flow analysis and visualization are based on a reliable and robust local rPPG signal. We apply the chrominance-based method [DJ13] to each spatial position to receive a local rPPG signal.

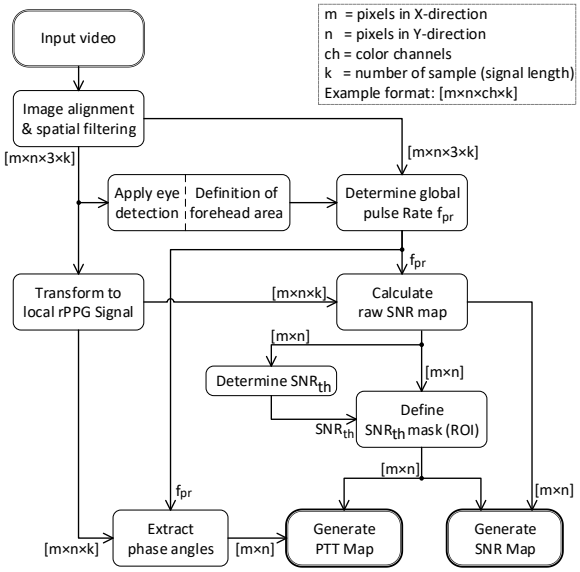

Figure 1: *Overall description of the framework to analyze and visualize the blood flow.*

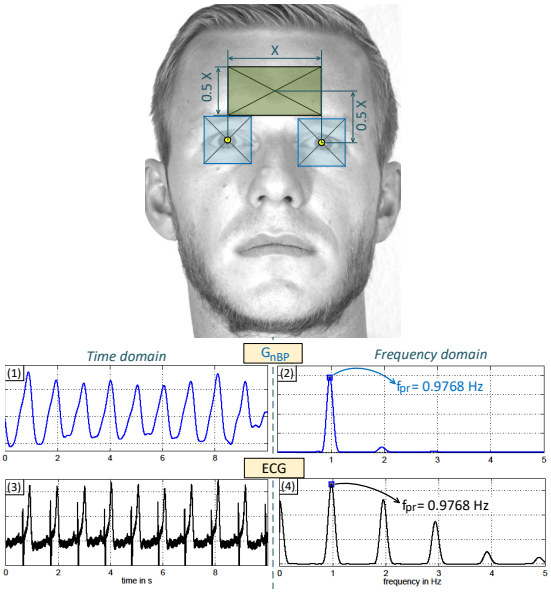

Figure 2: *The blue area represents the result of the used eye detector, and the green area is the derived forehead ROI. During the 10 s recording the subject has a pulse rate of $f_{pr} = 0.9768$ Hz, corresponding to 58.61 BPM*

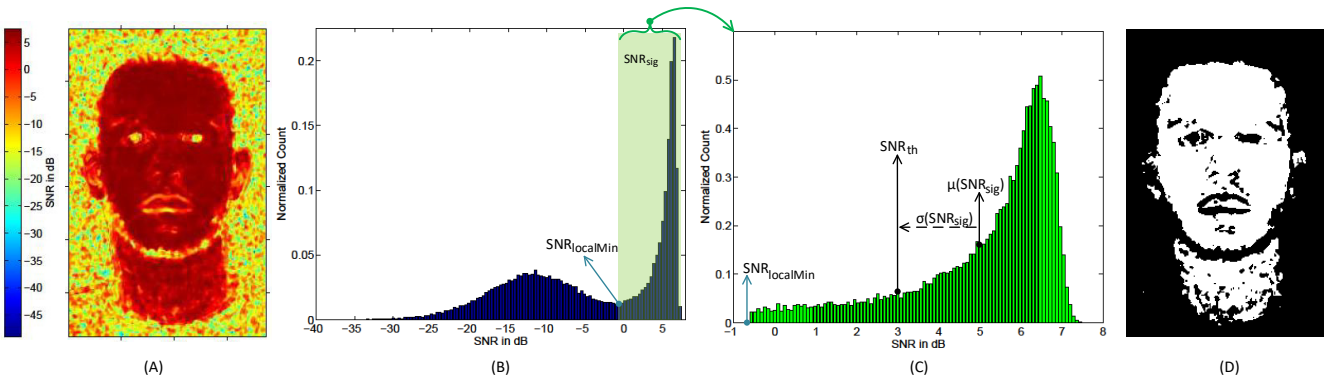

Figure 3: *This figure shows the separate steps of the SNR-threshold-mask generation. After spatial filtering (here average filter and double downsampling by a factor of two) the raw SNR map is calculated (A). From the resulting SNR values, we calculate a histogram and determine the local minimum ($SNR_{localMin}$) (B). All values below $SNR_{localMin}$ are discarded, and from the remaining histogram, the mean and standard deviation is calculated to determine SNR_{th} (C). The application of SNR_{th} leads to the SNR-threshold-mask, where all pixels above this value correspond to the visible skin tissue of the subject (D).*

2.2. SNR Map

We further use an SNR analysis to quantify the spatial signal quality. The SNR represents the strength of the signal of interest compared to the unwanted noise present in the signal. We use the SNR definition by de Haan and Jeanne [DJ13]. We calculate the SNR map from a video recording with a length of 10 s. During these 10 s, the subject's heart rate varies, which is referred to as heart rate variability (HRV) [SG17]. As in [ZTE*18], we consider the HRV by defining intervals of ±2 BPM around the fundamental frequency (f_{pr}) and the second harmonic as the signal. The remaining frequency components are classified as noise. As described in [dHvL14], the third and higher harmonics are not considered as a signal, because these frequency components are usually minimal, and the influence on the SNR is negligible. The SNR is calculated from the energy ratio around the pulse frequency f_{pr} plus the second harmonic $2f_{pr}$ and the remaining energy of the spectrum. Thus the SNR is defined by:

$$SNR = 10\log_{10}\left(\frac{\sum_{k=f_1}^{f_2}(U_m(k)X(k))^2}{\sum_{k=f_1}^{f_2}((1-U_m(k))X(k))^2}\right), \qquad (2)$$

where $X(k)$ is the magnitude of the signal of interest $x(t)$, f_1 and f_2 define the range in which the SNR is calculated (e.g., 0.5 Hz to 4.0 Hz), k is the bin number of the frequency component, and $U_m(k)$ is the binary mask. The SNR measures the energy ratio of the spectral segments inside (signal) and outside (noise) of the binary mask $U_m(k)$ which is described by:

$$U_m(k) = \begin{cases} 1, & if\ |f_{pr} - \Delta f \cdot k| \leq (2\,BPM) \\ 1, & if\ |2f_{pr} - \Delta f \cdot k| \leq (2\,BPM), \\ 0, & otherwise \end{cases} \qquad (3)$$

where Δf is the spectral frequency resolution. This resolution is calculated using:

$$\Delta f = \frac{f_s}{N_{fft}}, \qquad (4)$$

where N_{fft} is the number of FFT points (i.e., number of samples), and f_s is the sampling frequency of the input signal $x(t)$. The described SNR calculation is applied to each local rPPG signal and leads to a raw SNR map.

Our goal is to analyze only local positions that have an acceptable SNR. A corresponding segmentation is achieved by applying thresholding to the raw SNR map. Therefore, we calculate a binary mask (referred to as SNR-threshold-mask) containing only spatial positions that provide a reliable rPPG signal. The required SNR threshold SNR_{th} is calculated based on the statistical values of the raw SNR map. The determination of SNR_{th} is shown and analyzed in Section 2.3.

This analysis indicates that a sufficient segmentation of visible skin tissue is achieved with SNR_{th} and the resulting SNR-threshold-mask. All spatial positions above SNR_{th} represent the ROI, and the remaining positions are ignored for further processing. The first row of Figure 6 shows an SNR-threshold-mask for an image sequence and the resulting SNR map after applying this mask to the raw SNR map.

2.3. ROI Determination

In Section 2, we describe how to calculate an SNR map. These maps show the rPPG signal quality of the subject based on its current BVP. In the following, we describe how to calculate an SNR-threshold-mask, which then forms the ROI. This mask determines positions, where the subject's rPPG signal has reliable quality, and the mask can be used for segmentation purposes. By applying a certain threshold SNR_{th} to the raw SNR maps, we are able to segment the visible skin of a subject facing the camera.

We aim to determine a threshold that guarantees that all spatial positions above SNR_{th} are skin pixel. Based on the statistical values of the raw SNR map and its histogram (e.g., with 100 bins) the threshold SNR_{th} is defined. Figure 3 shows the steps to calculate the SNR-threshold-mask.

After spatial filtering, we calculate the SNR for each spatial position and map these values onto their corresponding position (A). As shown in Figure 3 (B), the SNR values are then plotted in the form of a histogram. This histogram has a local minimum ($SNR_{localMin}$) between two local maximums. The left maximum can be assigned to the background and the right maximum to visible skin tissue, which generally has a higher SNR. We split histogram (B) at the local minimum and focus on the SNR values larger than $SNR_{localMin}$. Plot (C) shows the remaining histogram. We calculate the statistical values of this histogram and set the threshold to:

$$SNR_{th} = \mu(SNR_{sig}) - \sigma(SNR_{sig}), \tag{5}$$

where SNR_{sig} represents the SNR values of the green-colored histogram, the operator σ corresponds to the standard deviation, and μ corresponds to the mean. All spatial positions with an SNR above SNR_{th} are classified as skin, and we omit the remaining positions. This thresholding leads to the SNR-threshold-mask shown in Figure 3 (D).

2.4. PTT Map

When the heart pumps blood volume into the aorta, it generates a pulse wave [BDJ*17]. This wave indicates the propagation and direction of the flowing blood volume. The PTT refers to the time difference between the pulse peaks taken at two arterial sites [iCM11].

In this work, we visualize the blood flow path through the face via PTT maps. We calculate the time difference (i.e., PTT) between signals of different position without any filtering via the phase angles in the frequency domain. Therefore, the FFT for each spatial rPPG signal is calculated. We then extract the phase angle $\varphi(f_{pr})$ of the frequency component f_{pr}. The phase difference of the spatial rPPG signals of interest are calculated by:

$$\Delta\varphi_i(f_{pr}) = \varphi_{ref}(f_{pr}) - \varphi_i(f_{pr}), \tag{6}$$

and can then be converted into a time delay by:

$$\Delta t_i = \frac{\Delta\varphi_i(f_{pr})}{2\pi f_{pr}}, \tag{7}$$

where $\varphi_{ref}(f_{pr})$ is the phase angle of a reference position, $\varphi_i(f_{pr})$ is the phase angle of a position of interest and f_{pr} is the global pulse frequency. This time difference Δt_i corresponds to the time that the BVP needs to travel from one point to the other. After all phase angles have been calculated, they are checked for plausibility, whereby angles that correspond to time delays of more than $\Delta t_{max} = 0.3\,s$ are excluded, as they are physiologically impossible [BDJ*17]. In order to remove these inaccurate phase angles, we calculate the mean of all phase angles within the ROI $\mu(\varphi_{sig})$ and remove all values outside the physiological range. Furthermore, we convert Δt_{max} with Equation (7) into a phase difference:

$$\Delta\varphi_{max}(f_{pr}) = \Delta t_{max} 2\pi f_{pr}. \tag{8}$$

The plausible PTT range is defined by:

$$\varphi(x,y) = \begin{cases} \varphi(x,y), & if\,[\varphi(x,y) > (\mu(\varphi_{sig}) - \frac{\Delta\varphi_{max}}{2})] \\ & and\,[\varphi(x,y) < (\mu(\varphi_{sig}) + \frac{\Delta\varphi_{max}}{2})], \\ NaN, & otherwise \end{cases} \tag{9}$$

where all phase angles φ correspond to the frequency component of f_{pr}. Subsequently, we select the spatial position with the largest phase angle as the reference point. This point corresponds to the position where the BVP appears last in time. In almost all frontal recordings, the position of the highest phase angle is located within or near the subject's mouth. To make the measurements of the individual subjects comparable, a reference point φ_{ref} of the PTT map is selected for all subjects within a rectangular area centered around the middle of the mouth. The phase differences $\Delta\varphi_i$ between the reference position and all other positions are calculated, and Equation (7) is applied to convert the resulting differences into the time delays Δt_i. Each time delay is color-coded (according to its value) and then mapped to the corresponding spatial position. Figure 5 shows PTT maps, where the BVP reaches blue areas first and red areas at a later point in time.

2.5. Data Acquisition

A dataset of 96 video sequences (duration 60 seconds each) of twelve healthy subjects between the ages of 25 and 33 years was captured. Each participant was connected to a vital sign monitor (VitaGuard 3100, GETEMED, Germany) to measure the ECG and PPG simultaneously to the video recording. In order to achieve higher robustness against motion artifacts, and due to the higher sampling frequency, we use the ECG measurements as reference data in this work. All sequences are recorded with a 4K resolution and 60 frames per second (fps) camera (PMW-F55 CinaAlta 4K, Sony Corporation, Japan). The recorded scene was illuminated by compact daylight (≈ 5600 K) LED sources. These light sources were placed in front of the subject with an angle of 30° to both sides, to reduce shadow on the subject's face. During the data acquisition, subjects were recorded in different positions (frontal and lateral), and with varying heart rate (resting and increased). The subjects increased their heart rate by performing bodyweight squats just before the start of the recording. All participants of the data acquisition provided their informed consent.

3. Results

3.1. SNR Map Evaluation

We evaluate the SNR maps, and SNR-threshold-mask of six randomly selected video sequences of different subjects to determine the effectiveness of the ROI determination defined from Section 2.3. We further compare the results for two different low-pass filters (Gaussian and average filters) in combination with a different number of downsampling steps (two and three times by the factor of two). In order to determine the skin segmentation quality of the SNR-threshold-mask, we manually define two different reference regions for each recording. The first region (see Figure 4 (B)) contains all head and neck pixels of the subject (see Figure 4 (A)). The second reference region contains only skin pixels (see Figure 4 (C)) without eyebrows, ears, mouth, eyes, thick beard, nostrils, and the transition from head to neck.

For each recording, we calculate the raw SNR map, determine the SNR_{th}, and generate an SNR-threshold-mask. We then multiply this mask with the manually selected reference regions and count

how many are outside the outline reference region (OR) and how many pixels are within the skin reference region (SR).

Table 1 lists the results for each recording. The percentage of pixels wrongly segmented outside the OR is always less than 1 % of all pixels for each recording and filter type. In the Gaussian filtered image sequences, it is problematic to determine a local minimum in the SNR histogram, which leads to an inaccurate calculation of the SNR_{th} and thus to a significant number of pixels, which are not classified as skin within the SR. For all image sequences processed by an average filter the number of correctly classified skin pixel within the SR is never below 79 % and mostly around 90 %.

Table 1: *Results for the tested image sequences and the analyzed spatial filtering techniques. It is listed how many percent of all pixels are wrongly or correctly segmented as skin.*

Subject	Filter - Level	SNR_{th} in [dB]	wrong detected outside OR [%]	correct detected inside SR [%]
ID01	Gauss-2	0.87	0.07	66.86
	Avg-2	3.08	0.01	93.73
	Gauss-3	1.67	0.03	93.63
	Avg-3	3.92	0.05	99.42
ID02	Gauss-2	0.13	0.18	37.64
	Avg-2	-0.40	0.25	89.05
	Gauss-3	0.07	0.24	79.03
	Avg-3	1.32	0.03	96.49
ID03	Gauss-2	1.28	0.07	23.16
	Avg-2	-0.67	0.16	88.85
	Gauss-3	-1.60	0.54	86.23
	Avg-3	1.05	0.10	92.92
ID04	Gauss-2	1.69	0.01	69.40
	Avg-2	2.70	0.01	96.24
	Gauss-3	1.51	0.01	96.63
	Avg-3	4.23	0.04	98.49
ID05	Gauss-2	0.03	0.12	82.78
	Avg-2	3.12	0.00	96.18
	Gauss-3	2.41	0.00	95.35
	Avg-3	3.83	0.00	98.45
ID06	Gauss-2	-2.44	0.46	39.41
	Avg-2	0.97	0.03	79.82
	Gauss-3	0.08	0.09	74.96
	Avg-3	2.37	0.02	92.09

3.2. Blood flow visualization

We assume that the time delay between rPPG signals of different regions corresponds to the time difference required by the peak of the BVP to reach these regions and thus corresponds to the PTT. Therefore, the visualization of the blood flow path is achieved with PTT maps, which map the calculated time delays to their spatial position. Figure 5 shows the resulting PTT maps for recordings of *ID01*, *ID02*, *ID03*, and *ID04*. These maps show similar results and patterns (see Figure 5) between each subject. Partial coverings of the skin tissue by head hair, beard, and jewelry lead to a weak SNR and cause that spatial positions are excluded. The SNR map visualizes the signal quality of each spatial position. The plots in Figure 5 show that the signal quality at the edges of the detected skin is rather low. Also, the overall SNR in the neck region is smaller than in the face.

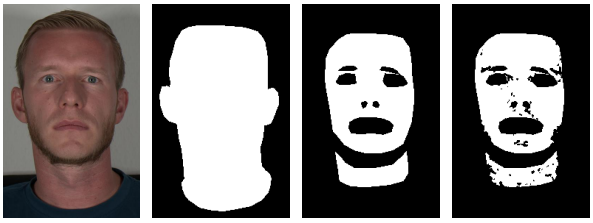

Figure 4: *This figure shows the input frame (A), the outline reference area (OR) (B), the skin reference area (SR) (C) and the multiplication result of SR and SNR-threshold-mask (D) for the average filtered and twice downsampled image sequence of ID01.*

The areas marked in blue in the PTT map represent certain areas which are reached early by the BVP. Since the blood flows via the common carotid into the head [BDJ*17], show the blue areas at the neck of the subjects, the position of the carotid artery. The forehead region of the head surface is supplied by the internal carotid arteries and the rest of the face by the external carotid arteries [vATYL18]. As expected, this difference in blood supply is also visible in the PTT maps. These maps show that the forehead is supplied earlier with blood than the rest of the face. The BVP reaches the cheeks and the mouth lastly.

3.3. Application in a PAD system

In this section, we present an experimental setup that demonstrates a direct application of our local blood flow analysis and visualization approach in a PAD system to detect persons wearing partial masks or heavy makeup. During the data acquisition, we recorded eight videos in which *ID01* wears different partial face masks. These recordings are used to demonstrate a presentation attack.

Figure 6 shows single frames of two PAD recordings in which the person wears a nose and chin mask. We process each input video, as shown in Figure 2 and calculate the SNR-threshold-mask, SNR map, and PTT map for each recording. The results with and without face coverage are shown in Figure 6. The individual plots in this figure show that recordings with a mask expose large excluded areas where the SNR is below SNR_{th}, or the PTT is outside the plausible PTT range.

A visual inspection of the PTT and SNR maps of Figure 6 indicates that the subject is wearing a nose mask in the second row and a chin mask in the third row. One simple way to automatically detect these large excluded areas could be to compare the generated maps with the results of a face detector. Therefore, we use the Viola and Jones eye and face detector [VJ01] to obtain the face and eye coordinates (illustrated by rectangles in Figure 6) of the subject. Based on these coordinates we define an ROI that extends over the y-coordinates of the face detector and the x-coordinate from the start of the left eye to the end of the right eye (see Figure 6). After

this ROI definition, we check how many pixels are excluded within this region.

For our experiments we defined that whenever more than 15 % of the SNR-threshold-mask is excluded, we classify that a presentation attack is present. Figure 6 shows in the first row that the person is not wearing a mask because only 9.11 % of the pixels are excluded, and therefore, no presentation attack is present. The second and third row of this figure exemplarily shows the results of two out of eight presentation attack videos. The two masks were recognized because the maps exclude more than 15 % of all pixels in the defined ROI. In this work, all eight masks could be classified as a presentation attack according to the described procedure.

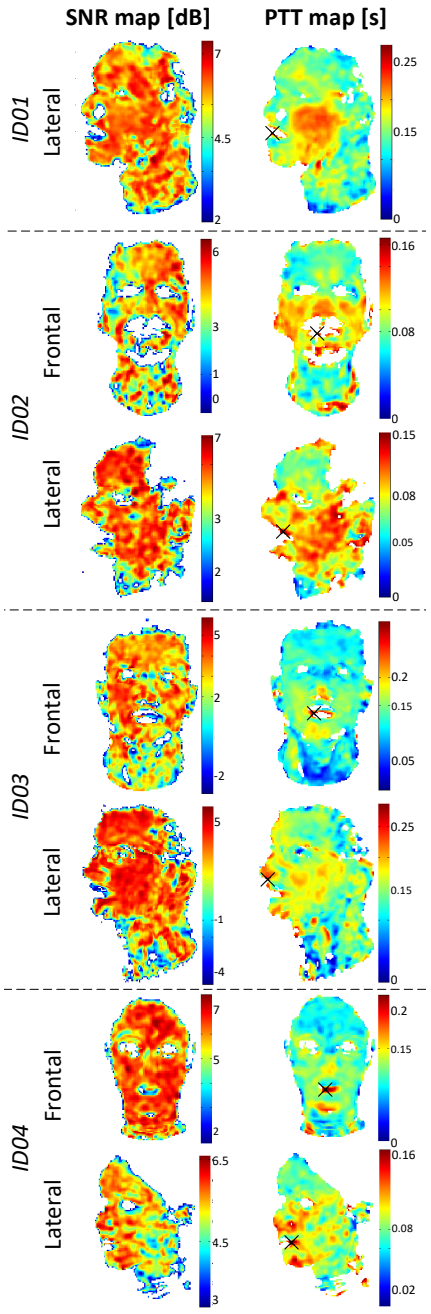

Figure 5: *Frontal and lateral maps of four different subjects (images are average filtered and three times downsampled). The black cross within the PTT maps indicates the position of* φ_{ref}.

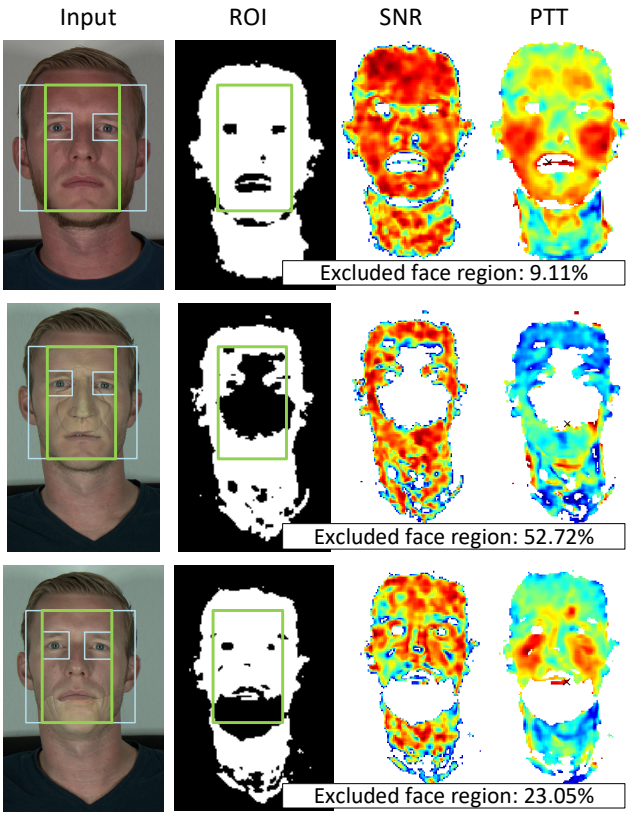

Figure 6: *Each row shows the results for a separate image sequence (filtered with an average filter and three times downsampled). The person wears a partial mask during the recordings in the second (nose mask) and third (chin mask) row. The blue rectangles within the input images illustrates the output of an eye and face detector. The green rectangle within the input anr ROI images shows the calculated ROI used to automatically detect a presentation attack.*

4. Conclusion

In this paper, we propose a novel approach to analyze and visualize the local blood flow based on a chrominance-based rPPG signal. The visualization of the local blood flow analysis clearly shows blood flow paths known from literature (e.g., the position of the common carotid arteries), as shown in Figure 5.

Additionally, we can segment visible skin tissue based on the pulse rate of the subject. In our investigation, the spatial filtering of the input images with an average filter led to the best results. With the proposed SNR_{th} calculation an accurate and reliable skin segmentation is achieved.

This skin segmentation and the blood flow analysis can be used for liveliness detection against presentation attacks. We tested the application for PAD with recordings where parts of the face are covered. All recordings are correctly classified as a presentation attack. This approach shows promising results and could be used in automated border control systems at airports.

Other potential applications include physiological measurements in medical applications, e.g., intraoperative blood flow visualization. It is also conceivable that the presented analysis and visualization can be used to differentiate between different soft tissues during surgery similarly as in [WUA*18, WKU*19, WRE*19].

Acknowledgment

The work in this paper has been funded in part by the German Federal Ministry of Education and Research (BMBF) through the projects 3DFinder under the grant number 03ZZ0445F and MultiARC under the grant number 16SV8061.

References

[BDJ*17] BETTS J. G., DESAIX P., JOHNSON J. E., KOROL O., KRUSE D., POE B., WISE J., WOMBLE M. D., YOUNG K. A.: *Anatomy & Physiology*. Open Textbooks. OpenStax College, Rice University, 2017. URL: https://cnx.org/contents/FPtK1zmh@11.1:fEI3C8Ot@14/Preface. 5, 6

[dHvL14] DE HAAN G., VAN LEEST A.: Improved motion robustness of remote-PPG by using the blood volume pulse signature. *Physiological Measurement 35*, 9 (sep 2014), 1913–1926. doi:10.1088/0967-3334/35/9/1913. 4

[DJ13] DE HAAN G., JEANNE V.: Robust pulse rate from chrominance-based rPPG. *IEEE Transactions on Biomedical Engineering 60*, 10 (2013), 2878–2886. doi:10.1109/TBME.2013.2266196. 1, 2, 3, 4

[GMF14] GALBALLY J., MARCEL S., FIERREZ J.: Biometric antispoofing methods: A survey in face recognition. *IEEE Access 2* (2014), 1530–1552. doi:10.1109/ACCESS.2014.2381273. 2

[HM19] HEUSCH G., MARCEL S.: Pulse-based features for face presentation attack detection. *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems, BTAS 2018* (2019). doi:10.1109/BTAS.2018.8698579. 1, 2

[iCM11] I CARÓS S., MARIA J.: *Continuous non-invasive blood pressure estimation*. PhD thesis, ETH Zurich, 2011. 5

[KKLP15] KWON S., KIM J., LEE D., PARK K.: ROI analysis for remote photoplethysmography on facial video. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2015-November* (2015), 4938–4941. doi:10.1109/EMBC.2015.7319499. 3

[LKZ*17] LI X., KOMULAINEN J., ZHAO G., YUEN P. C., PIETIKAINEN M.: Generalized face anti-spoofing by detecting pulse from face videos. *Proceedings - International Conference on Pattern Recognition* (2017), 4244–4249. doi:10.1109/ICPR.2016.7900300. 1, 2

[MSC16] MARASCO E., SHEHAB M., CUKIC B.: A methodology for prevention of biometric presentation attacks. *Proceedings - 7th Latin-American Symposium on Dependable Computing, LADC 2016* (2016), 9–14. doi:10.1109/LADC.2016.13. 2

[NSV17] NOWARA E. M., SABHARWAL A., VEERARAGHAVAN A.: PPGSecure: Biometric Presentation Attack Detection Using Photoplethysmograms. *Proceedings - 12th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2017 77005* (2017), 56–62. doi:10.1109/FG.2017.16. 1, 2

[PMP10] POH M.-Z., MCDUFF D. J., PICARD R. W.: Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics Express 18*, 10 (2010), 10762. doi:10.1364/OE.18.010762. 1, 2

[RWAH17] RAPCZYNSKI M., WERNER P., AL-HAMADI A.: Continuous low latency heart rate estimation from painful faces in real time. *Proceedings - International Conference on Pattern Recognition* (2017), 1165–1170. doi:10.1109/ICPR.2016.7899794. 1, 2, 3

[SG17] SHAFFER F., GINSBERG J. P.: An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health 5*, September (2017), 1–17. doi:10.3389/fpubh.2017.00258. 4

[SL16] SUH K. H., LEE E. C.: Face liveness detection for face recognition based on cardiac features of skin color image. *First International Workshop on Pattern Recognition 10011*, July 2016 (2016), 100110C. doi:10.1117/12.2242472. 1, 2

[TL15] TSOURI G. R., LI Z.: On the benefits of alternative color spaces for noncontact heart rate measurements using standard red-green-blue cameras. *Journal of Biomedical Optics 20*, 4 (2015), 048002. doi:10.1117/1.JBO.20.4.048002. 1, 2

[TMSY14] TAMURA T., MAEDA Y., SEKINE M., YOSHIDA M.: Wearable Photoplethysmographic Sensors—Past and Present. *Electronics 3*, 2 (2014), 282–302. doi:10.3390/electronics3020282. 1

[vATYL18] VON ARX T. V. A., TAMURA K., YUKIYA O., LOZANOFF S.: The Face – A Vascular Perspective. A literature review. *Swiss dental journal 128*, 5 (2018), 382–392. 6

[VJ01] VIOLA P., JONES M.: Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001 1* (2001), I–511–I–518. doi:10.1109/CVPR.2001.990517. 3, 6

[VSN08] VERKRUYSSE W., SVAASAND L. O., NELSON J. S.: Remote plethysmographic imaging using ambient light. 72–75. arXiv:NIHMS150003, doi:10.1364/OE.16.021434. 3

[WKU*19] WISOTZKY E. L., KOSSACK B., UECKER F. C., ARENS P., DOMMERICH S., HILSMANN A., EISERT P.: Validation of two techniques for intraoperative hyperspectral human tissue determination. In *Proceedings of SPIE* (2019), vol. 10951, p. 109511Z. doi:10.1117/12.2512811. 8

[WRE*19] WISOTZKY E. L., ROSENTHAL J., EISERT P., HILSMANN A., SCHMID F., BAUER M., SCHNEIDER A., UECKER F. C.: Interactive and Multimodal-based Augmented Reality for Remote Assistance using a Digital Surgical Microscope. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (mar 2019), pp. 1477–1484. doi:10.1109/VR.2019.8797682. 8

[WRS*12] WU H.-Y., RUBINSTEIN M., SHIH E., GUTTAG J., DURAND F., FREEMAN W.: Eulerian video magnification for revealing subtle changes in the world. *ACM Transactions on Graphics* (2012). doi:10.1145/2185520.2335416. 1, 3

[WUA*18] WISOTZKY E. L., UECKER F. C., ARENS P., DOMMERICH S., HILSMANN A., EISERT P.: Intraoperative hyperspectral determination of human tissue properties. *J. Biomed. Opt. 23*, September (2018), 1–8. doi:10.1117/1.JBO.23.9.091409. 8

[Yan15] YANG JUN, GUTHIER BENJAMIN E. S. A.: Estimating two-dimensional blood flow velocities from videos. *International Conference on Image Processing (ICIP)* (2015), 3768–3772. 2

[ZTE*18] ZAUNSEDER S., TRUMPP A., ERNST H., FÖRSTER M., MALBERG H.: Spatio-temporal analysis of blood perfusion by imaging photoplethysmography. *Optical Diagnostics and Sensing XVIII: Toward Point-of-Care Diagnostics 10501*, February (2018), 32. doi:10.1117/12.2289896. 1, 2, 4

[ZTWM18] ZAUNSEDER S., TRUMPP A., WEDEKIND D., MALBERG H.: Cardiovascular assessment by imaging photoplethysmography - A review. *Biomedizinische Technik* (2018), 1–18. doi:10.1515/bmt-2017-0119. 1